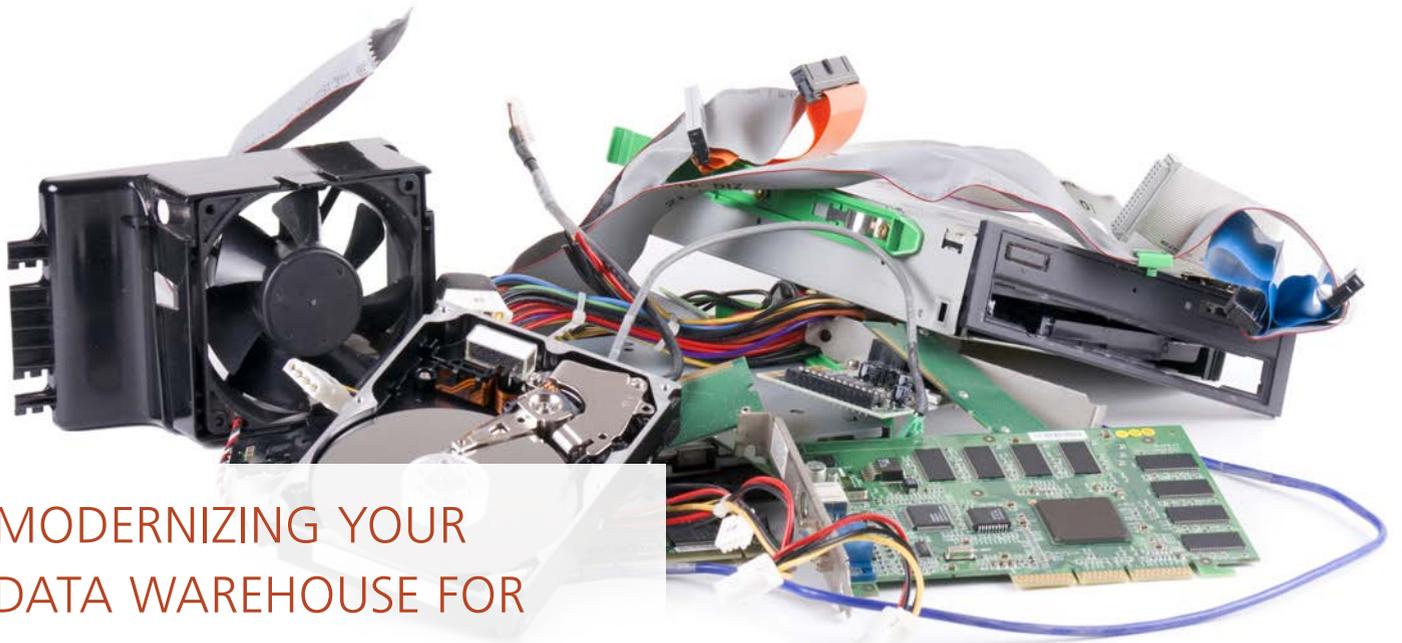


DATA INNOVATIONS

SHOWCASE

Modernizing Your Data Warehouse for Cutting-Edge Analytics

A CONVERSATION WITH HPE VERTICA



MODERNIZING YOUR DATA WAREHOUSE FOR CUTTING-EDGE ANALYTICS

Updating your data warehouse to handle more data and complex analytics need not require major disruptions to your environment. With its combination of speed and MPP scalability, a columnar analytic database such as HPE Vertica has a critical role to play in the hybrid big data architecture of tomorrow—and it's economical to boot.

Is your aging data warehouse system running out of gas? Can it no longer keep pace with existing workloads? Is it unfit for new workloads?

If so, you might be tempted to pull the plug on it altogether, reasoning that the data warehouse is a vestige of the distant past and that data warehouse systems don't have a role to play in big data architecture. You wouldn't just be making a mistake—you'd be making an *extremely expensive* mistake.

Your aging data warehouse is the problem, but a data warehouse is still the best solution. It always has been. Pound for proverbial pound, the data warehouse

remains a query-processing platform par excellence. Data warehouse architecture is faster *and* cheaper—much, much cheaper—than any other alternative, including new, highly hyped platforms such as Hadoop or Apache Spark. Both Hadoop and Spark complement and extend the data warehouse in critical ways. They likewise address use cases or workloads for which the data warehouse was not designed. However, Hadoop, Spark, and other platforms are ill-suited for analytical query processing, which is the warehouse's *raison d'être*.

That said, much has changed since data warehouse architecture was first codified 25 years ago. The row-based, non-parallel-processing databases that served as the bedrock of traditional data warehouse systems have been superseded by much faster, and considerably more efficient, columnar data warehouses. These columnar databases, which store data in columns instead of columns and rows, make use of a technology called massively parallel processing, or MPP.

MPP isn't a new thing, but its democratization is. A decade ago, analytic databases such as Vertica from Hewlett Packard Enterprise (HPE) radically revolutionized the economics of MPP, delivering massive

scalability at affordable price points. For the vast majority of workloads, an MPP data warehouse will perform faster—sometimes twice as fast (or more)—than will a non-parallel-processing RDBMS. For the majority of *analytic* workloads, a columnar MPP data warehouse will be even faster.

That aging data warehouse that's costing you a premium to maintain and no longer can return query answers quickly? It's almost certainly powered by a row-based, non-parallel-processing database. You're just asking too much of it. Replacing it with an analytic data warehouse such as HPE Vertica will supercharge performance for your existing applications, permit you to develop challenging new analytics, and, most compelling of all, provide a logical, predictable, and affordable scale-out strategy.

This is because MPP data warehouses scale (mostly) linearly: if you double the footprint of a two-node MPP warehouse, you effectively double its performance. Because of the way an MPP analytic database such as HPE Vertica is priced—i.e., on a per-terabyte (TB) basis—it's possible to scale data warehouse performance linearly, in response to business needs, in a transparent manner. This is a radical departure from traditional per-processor, per-node, per-concurrent-user, per-connected-system—or per-all-of-the-above—MPP pricing schemes.

“Companies are paying too much for their legacy systems. They're paying licensing fees for the number of users, the number of connected systems, and so on. These old licensing schemes are dinosaurs. They're costing them a ton of money, and on top of that, they're predicated on the use of this legacy technology that really wasn't meant to handle big data,” argues Steve Sarsfield, product marketing manager with HPE Vertica.

“A platform like HPE Vertica solves both of those problems. It can scale to handle even the most extreme performance requirements. It has the advantage of a very transparent, per-terabyte pricing model. This translates into a simple way to scale out. Customers can easily modernize and reduce the amount of spend to deliver cutting-edge analytics.”

The Crux of Columnar

Data warehouse architecture developed in a context in which row-store online transaction processing (OLTP) relational databases such as Oracle, DB2, and Sybase predominated.

These OLTP RDBMSs ran on both mainframe kit (in the case of Oracle and DB2) and on client-server systems—i.e., platforms such as Netware, OS/2, Unix, and (later) Windows NT. For this reason, OLTP databases tended to be cheaper and more affordable than predecessor platforms such as IMS and ADABAS, which ran on costly mainframe hardware, and which—thanks to Big Iron's complex capacity-pricing schemes—were prohibitively expensive to maintain.

Analytics workloads typically involve only a select few attributes, which means that—for most kinds of analytics—a columnar design will offer superior performance and efficiency.

In other words, the row-store database became the de facto foundation for the data warehouse because it was the best—i.e., the most practical, cost-effective, and portable—solution available at the time. This is no longer the case, and, in fact, hasn't been for a while, thanks to the development and maturation of columnar database solutions.

Unlike OLTP workloads, analytics workloads typically involve only a select few attributes, which means that—for most kinds of analytics—a columnar design will offer superior performance and efficiency. When the first data warehouse systems were developed and implemented, however, there were no viable commercial columnar databases. Some platforms, such as IMS, offered columnar-like capabilities, but weren't based on the relational model. (IMS was also a mainframe-only proposition, which doubly ruled it out.)

HPE Vertica was one of the first of the so-called “analytic”—or columnar MPP—data warehouse systems. It was conceived by database design guru

Michael Stonebraker, the brains behind the seminal Ingres RDBMS and its successor, Postgres, as well as VoltDB, “Tamr,” and a slew of other cutting-edge database or data management systems. Stonebraker reasoned that the row-store database had basically been “grandfathered in” as a foundation for the data warehouse. He knew that a columnar store would deliver significantly improved analytical performance, along with several other critical advantages.

A columnar design helps to minimize I/O contention—by far the biggest bottleneck in analytical processing—as well as to drastically reduce disk seek times.

“The magic of a columnar database is its superior performance. In traditional relational databases, data is stored in rows, which means that even if you have a query that only needs data from a single column, the database goes in and scans the contents of all of those rows—every single column in every single row. The entire row has to be read from the disk,” Sarsfield explains. For most types of analytics, he notes, you’re primarily interested in a single column’s worth of attributes. The rest, so to speak, is noise. “That’s pretty high on I/O, and it’s pretty high on seek time and latency, which means it’s pretty high on the amount of data the database has to read in order to answer a simple query.”

A columnar design helps to minimize I/O contention—by far the biggest bottleneck in analytical processing—as well as to drastically reduce disk seek times. These are two of columnar’s three biggest benefits. The other is its extremely high “compressability,” which exceeds that of a row-based database by a factor of 4 or 5.

“Columnar designs already compress very well just by virtue of the architecture itself, and HPE Vertica offers over a dozen different schemes and algorithms to help compress the data. HPE Vertica is able to look specifically at the data itself and then to make a decision as to which compression algorithm to use,” he explains, noting that HPE Vertica’s columnar design permits the use of late materialization, which means it can perform operations on compressed columns in-memory.

This likewise boosts performance—in some cases, speeding up performance by a factor of 3—if the compressed data does not need to be combined with data from one or more other compressed columns. (In such cases, all applicable columns must be uncompressed before data can be operated on or combined.)

“The ratio of compressability that we’re able to achieve gives us, comparatively, a much smaller footprint in terms of the amount of space that you actually need to store data. It also gives us even better I/O performance,” Sarsfield says. “For example, when I have date fields, HPE Vertica will automatically use the best algorithm for compressing dates, so I’m using the least amount of space.”

Row-store RDBMSs, by contrast, do not compress nearly so well. What’s more, in practice, row-store databases that are pressed into service as data warehouse systems are held together by the equivalent of decision-support duct tape. They require considerably more tuning or balancing—especially as data volumes increase—than do columnar databases. In addition, DBAs must implement workarounds (materializing views, vertically partitioning schema, or indexing every column) to maintain performance. These workarounds aren’t kludgy in and of themselves, but they add to the cost of operating and maintaining the data warehouse system, and as data volumes and query types and complexity increase over time, their effectiveness likewise degrades.

“I’ve seen hundreds or thousands of rows in a database, and you can imagine how slow that is when you have to read in all of them. You’d typically implement one or more workarounds—things like materialized views or indices—and in such cases, the best you can hope for is that you can try to get your queries answered in a ‘reasonable’ amount of time, but reasonable doesn’t cut it,” Sarsfield points out. “These workarounds have limitations. Materialized views have to be updated and maintained, indices have to be pre-built, and so on. Ultimately, they can accomplish just so much.”

Columnar isn’t a silver bullet. No technology is or could be. Its few shortcomings are comparatively mild, however. More to the point, columnar’s drawbacks can be more than offset by a combination of MPP horsepower and database design tweaks. One such

shortcoming, for example, is that a columnar database stores each column in its own set of disk blocks. (The columnar database, like its relational counterpart, optimizes for physical storage on-disk or in-memory. This is a critical difference between a database management system, such as a columnar or OLTP RDBMS, and a data store, such as Hadoop. We'll have more to say about this in the next section.)

If a columnar design is ideal for analytical processing, massively parallel processing is ideal for scaling analytical processing.

In a vanilla columnar implementation, this can result in slower inserts and updates, as well as a protracted—or more complicated—loading process. Columnar designs such as HPE Vertica control for this by implementing optimized, massively parallel loader technologies and by making use of relational OLAP (ROLAP) or multidimensional OLAP (MOLAP) techniques. Plus, Sarsfield notes, a columnar design permits the use of other optimizations such as block iteration that can likewise accelerate performance. For example, multiple values from a column can be passed as a block from one operator to the next. This is more efficient than in a row-oriented scheme, which uses per-tuple iterators.

"We have very large customers, such as Facebook, that load over 35 TB an hour into HPE Vertica. They obviously have some special things that they do to achieve that. In Facebook's case, they believe that they can actually load up to 60 TB an hour, if needed, but 35 TB is what they need now. For requirements like that, we can set aside nodes that are specifically assigned for just loading data, as well as set aside nodes for just performing analytics," he explains. "We've also had requests from our customers to have this kind of ROLAP/MOLAP functions available as part of our database, so that is now another thing that you can do with the HPE Vertica platform. That's a fairly new feature."

Currently, several vendors implement column-store-like capabilities in row-oriented database designs. For example, Microsoft's forthcoming SQL Server 2016 database will support "column-store indexing," and its PowerPivot Excel add-in enables the equivalent of

a column store on the desktop. Not only aren't these offerings true columnar database designs, but, in most cases, they aren't MPP databases, either.

If a columnar design is ideal for analytical processing, massively parallel processing is ideal for scaling analytical processing. Sarsfield cites the experiences of HPE Vertica customers, which he says illustrate the compelling—and sometimes incredible—difference of columnar MPP performance.

"A year or so ago we did a survey of our customers who were looking at the difference in query performance between their old data warehouse systems and their new HPE Vertica warehouses. About 10 percent of customers said they got as much as a 1,000 percent boost in performance, but 64 percent of our customers said they got over a 100-percent performance boost," he says. "That 1,000 percent boost means ... that what used to take an hour now takes just six minutes, or what used to take eight hours now takes just 48 minutes."

Harnessing Hadoop

Hadoop is a massively parallel processing environment, too. Unlike an RDBMS, however, Hadoop is not a database management system. Its storage layer is a file system, the Hadoop Distributed File System (HDFS), which stores and manages data across a cluster-wide configuration of nodes.

An MPP RDBMS such as HPE Vertica distributes and manages data across a cluster of nodes, too. The primary difference between an RDBMS and a file system—which, after all, is a database of a sort—is a layer of abstraction. In most cases, a database management system works by superimposing physical structure and placement optimization, logic, and schema on top of a file system substrate. HDFS, by contrast, has no native facility for doing this. Like an RDBMS, Hadoop can efficiently store and retrieve data, which it records in "blocks" of a consistent size. Unlike an RDBMS, it has no built-in facility for relating the data that's stored in these blocks.

If Hadoop is to behave like a SQL query platform, users must instead implement add-on technologies, such as Hive, Impala, Drill, or Spark (with Spark SQL) to replicate some of the functions of an RDBMS. From

a DM perspective, however, all of these tools are impoverished. Unlike HPE Vertica, for example (which complies with the latest (SQL 2011) version of the ANSI SQL Standard), Hive, Impala, and Spark SQL are not fully ANSI SQL-compliant. Hive, Impala, and other Hadoop SQL query engines also lack strong schema support and—relative to an optimized platform such as HPE Vertica—don't support high concurrency levels.

Hadoop can be used to complement, to extend, and even to outstrip the data warehouse in important ways.

"A lot of folks are looking to Hadoop to solve the problems they're having with their legacy data warehouse systems, and it's true that the legacy architectures can't handle a lot of the analytical workloads people are developing. If you compare Hadoop's performance for ad hoc query and SQL analytics versus HPE Vertica's, it isn't even close: HPE Vertica is an order of magnitude or faster for most queries," Sarsfield claims.

"In order to support data warehousing workloads on Hadoop, you have to invest in and develop new skills, new software, and new people. What we're offering with HPE Vertica is a way out of the costly legacy data warehouse paradigm. We're offering a way to modernize without major disruption."

HPE's strategy is in no sense anti-Hadoop, stresses Sarsfield, who notes that Hadoop can be used to complement, to extend, and even to outstrip the data warehouse in important ways.

First, because Hadoop runs on top of a distributed file system, it can efficiently ingest data of virtually any kind or type. An MPP RDBMS, by contrast, imposes strict constraints on the kinds or types of data it's able to ingest. (Data must conform to a schema, and this schema must—with certain exceptions—be defined in advance.) An MPP RDBMS is likewise at a disadvantage when it comes to processing heterogeneous data types: relational database systems are optimized for query processing, which involves operations (using the SQL language) on structured, tabular data. As we've noted, it's possible to cost-effectively land, store, and process

structured, semi-structured, and multi-structured data in Hadoop. This simply isn't the case with a relational database.

If anything, Sarsfield claims, HPE gives customers the best of both worlds with HPE Vertica. On the one hand, they're able to leverage Hadoop as a cost-effective platform for persistence and light data management. On the other hand, they can use HPE Vertica to accelerate both traditional data warehouse workloads and advanced analytics.

"If you want to store data away for a rainy day, if you're not really sure what value it has—maybe it's Weblogs, maybe it's data that's coming in from the Internet of things—if you want to do data discovery on that, HPE Vertica is great for that. Really, if you need analytics that is very fast on massive amounts of data regardless of the source, HPE Vertica is ideal," he says, citing the example of online coupon aggregator RetailMeNot that uses HPE Vertica analytics to serve up timely coupons to both online and brick-and-mortar subscribers.

HPE engineered HPE Vertica so that it can actually query directly against data stored in Hadoop.

"They have a lot of analysis on who clicks through and uses their coupons versus who doesn't use them," he reveals. "All of that stuff is dumped into Hadoop where it's stored, but once they want to deliver that to their customers, they use HPE Vertica to actually process that analytical information."

A Hybrid Architecture for Big Data

The upshot is that there's no zero-sum choice between a data warehouse or Hadoop, Sarsfield argues. Instead, it's an issue of both-and. Companies such as Facebook and RetailMeNot implement both an HPE Vertica data warehouse and Hadoop in the context of a hybrid architecture for big data.

For these and other like-minded companies, Hadoop provides an inexpensive and scalable data storage substrate. Industry luminary Thomas Davenport, a senior research fellow at MIT's Center for Digital Business, puts the cost of Hadoop-based storage at \$ 0.23 per gigabyte, compared to \$19 per gigabyte in a traditional

data warehouse. Even though the per-gigabyte cost of storage in a columnar MPP platform such as HPE Vertica is much cheaper than that of a traditional MPP or non-MPP data warehouse, Hadoop is still a more flexible data storage option.

HPE Vertica gives customers the best of both worlds: they can leverage Hadoop as a cost-effective platform for persistence and light data management as well as accelerate traditional data warehouse workloads and advanced analytics.

Facebook, RetailMeNot, and others aren't using Hadoop to replace their legacy data warehouse systems, however; instead, they're leveraging columnar MPP platforms such as HPE Vertica to modernize these data warehouse investments. HPE Vertica gives them a cost-effective and scalable analytics engine for both traditional and advanced analytics, as well as an extensible data management platform that can help them get the most out of what they're doing with Hadoop.

The future is about leveraging the strengths of each platform how, when, and where it makes sense to do so, Sarsfield argues. This is one reason HPE engineered HPE Vertica so that it can actually query directly against data stored in Hadoop.

"We knew that this would be helpful because a lot of our customers have Hadoop nodes that they've sort of experimented with. They've set up sandboxes in Hadoop clusters and they've put data on them, but once they try to do analysis on what they have, they find that there are some limitations in terms of what's practicable with Hadoop. They can't do all of the analytical functions they can do with an HPE Vertica data warehouse. Hadoop doesn't have rich metadata catalog services. Hadoop can't support extremely high concurrency levels," he points out, noting that the HPE Vertica engine can also operate on Hadoop's equivalent of columnar-like storage—namely, Parquet and Optimized Row Columnar (ORC) files, which are used with Impala and Hive, respectively.

"We've taken our own SQL query engine that runs on HPE Vertica and we've opened that up so that you can use it to access data that's sitting directly in Hadoop. It can be a Parquet file or an ORC file—it doesn't matter. We can access files that are sitting in Hadoop and can perform analytics on them as well as using some of the other optimized formats that they have."

This is the logic behind HPE Haven, a hybrid big data platform that consists of HPE Vertica, HPE's Hadoop-based Intelligent Data Operating Layer (IDOL), and HPE's Distributed R technology. HPE IDOL is an engine for text analytics and for the analysis of multi-structured data, such as audio, video, and images, as well as files of any kind. Distributed R is a massively parallel implementation of R, the open-source statistics programming environment. HPE developed Distributed R in order to parallelize R processing across multiple nodes.

Facebook, RetailMeNot, and others are leveraging columnar MPP platforms such as HPE Vertica to modernize their data warehouse investments.

"The idea behind HPE Haven is that we're trying to provide an integrated platform for analytics on data of all shapes and sizes. First, there's structured data, such as OLTP or CRM data. Second, there's machine data—IoT, ATMs, all sorts of machines, including server logs and the like. Then there's this third stream, which we call 'human data'—voice, video, audio, facial and license plate recognition, stuff like that. The Haven platform has three best-of-breed engines that are optimized for each of these very different kinds of processing. At its core, it uses the HPE Vertica engine to handle the structured data, so Vertica gives HPE Haven a highly scalable ANSI-SQL analytical engine," Sarsfield explains.

"HPE IDOL provides a great way to understand that human information: the social data and also the voice, the video, and the multi-structured data. Then we also have Distributed R, which is the predictive analytics engine we're using. The three engines work together to really provide a great way to do standard analytics,

advanced analytics, and predictive analytics—and to do them all at scale, at the speed you need.”

Don't think of HPE Haven as a kitchen-sink approach to analytical processing, Sarsfield cautions. Think of it, instead, as a pragmatic “vision of how to leverage the strengths of each best-of-breed analytical engine—of how to tie it all together.” To this end, he explains, “There are hooks between all of Haven's engines, so if you want to use Distributed R and HPE Vertica together, you can use the optimizations that exist in HPE Vertica to speed up your predictive analytics. You can use HPE Vertica to automatically load data in parallel or to schedule tasks for parallel execution in Distributed R.”

No One-Size-Fits-All Solution

No single platform is ideal for any or every possible analytical workload. HPE's Vertica database is a peerless platform for query processing and for advanced analytics on structured or mostly structured data. It would be ill-used as an engine for analyzing multi-structured image or video files, however.

“Unlike Hadoop, HPE Vertica was really built for these kinds of analytical workloads. If you look at Hadoop, it was built for distributing file-based storage over a cluster and distributing MapReduce over a cluster,” Sarsfield says.

“The kinds of advanced analytics you'd do in HPE Vertica are unlike those you'd do in Hadoop. Hadoop is ideal for analytics on certain kinds of multi-structured data. They're both highly scalable platforms, they're just used for different purposes.”



Deriving greater value from your Enterprise Data Warehouse

Drive down cost with the HPE Vertica Analytics Platform



Figure 1: Teradata or Oracle vs. HPE Vertica Analytics Platform

HPE Vertica benefits

Skip costly upgrades

The HPE Vertica Analytics Platform is a software-only SQL analytics solution.

Increase team productivity

Operations that took days, now take hours and hours now take seconds.

Be prepared for massive scale

Infinitely and easily scale your Big Data analytics solution.

Data warehouses are extremely large, exorbitantly expensive legacy software architectures tied to larger, costly hardware architectures built to accommodate large volumes of data.

Big Data, however, is threatening the very existence of the data warehouse—which is bursting at the seams just to contain this data. More importantly, the performance of traditional data warehouse architectures is rapidly degrading under pressure from the Volume, Velocity, and Variety of Big Data. The only way to keep up is to throw more money (a lot more money) at the warehouse operations—but even that, at some point, can run out of runway.

Warehouses have walls—but your Big Data Analytics Platform can't

Data today is external as much as it is internal—sensor, customer, operations, patient, sentiment, and a variety of other

data sources. This semi-structured and unstructured data simply cannot be stored in traditional data warehouses.

A “warehouse” implies that your data is stored away, but not easily accessible. This outdated approach is focused on keeping the data rather than using the data—and the bigger the warehouse, the dustier the corners become and the more challenging it becomes to find the data when you actually need to get answers.

No walls, No boundaries

So, you have made a major investment in your data warehouse. However, squeezing out marginal performance gains requires you to constantly invest in more hardware and services. There has to be a better, more cost-effective Big Data SQL analytics solution.

Real-world, bottom line results

COMPANY	WITH TRADITIONAL DATA WAREHOUSE	WITH HPE VERTICA ANALYTICS PLATFORM	BUSINESS BENEFIT ACHIEVED
Leading online gaming company	<ul style="list-style-type: none"> • Could not achieve business-critical objectives 	<ul style="list-style-type: none"> • Analyzing 70 billion rows per day in real time • Running 100x faster than other data warehouse in consideration • Adding 8 TB of raw data daily on a 230-node cluster • Store 2 PB of source data compressed to 500 TB • Generate 500–600 unique reports per day • Run 10,000 total reports per day 	<ul style="list-style-type: none"> • Saved \$30 million USD by choosing HPE Vertica Analytics Platform instead of data warehouse solution vendor
Cardlytics	<ul style="list-style-type: none"> • Up to 20 hours to return a query result 	<ul style="list-style-type: none"> • 4–80x faster (return queries as fast as 30 seconds) • 5 percent overall cost of legacy platform • 90 percent reduction in operational overhead (and can reassign personnel from maintenance to revenue-focused analytics) 	<ul style="list-style-type: none"> • Achieved an ROI in three months • Increased average customer pipeline 10x • 200 new merchant prospects on a weekly basis, up from 20 per week on legacy platform
Global telecommunications company	<ul style="list-style-type: none"> • Processes took 14 hours • Cost \$120,000 USD per TB on license costs alone 	<ul style="list-style-type: none"> • Processes take 1 hour • Cost less than \$5,000 USD per TB with license and hardware costs 	<ul style="list-style-type: none"> • Save hundreds of thousands of dollars in license and hardware costs
Large U.S. bank	<ul style="list-style-type: none"> • Could not handle real-time scoring of credit cards (approvals or denials of credit cards) 	<ul style="list-style-type: none"> • 4-node cluster 80 percent faster than a 12-node cluster at one-third of the cost 	<ul style="list-style-type: none"> • Saved \$20 million USD in license costs on one initiative • Saved \$4 million USD in license and hardware costs on other initiative • Now meets SLAs

Increasingly, organizations are supplementing or replacing their enterprise data warehouse solutions with the HPE Vertica Analytics Platform for these reasons:

- **Save millions of dollars in license and hardware costs**—With optimized data storage, you can store 10–30x more data per server than traditional data warehouses and save even more with a straightforward and affordable pricing model.
- **Improve performance by orders of magnitude**—Gain insights into your data in near-real time by running queries 50x–1,000x faster than legacy data warehouse solutions.
- **Skip the costly and cumbersome “refrigerator” upgrades**—The HPE Vertica Analytics Platform is a software-only solution, so you have the freedom to choose the hardware that suits your environment, and you can load and query data concurrently so you do not need to take your system offline for upgrades.

- **Increase team productivity and organizational value**—Because operations that took days now take hours and hours now take seconds, your analytics team can be more productive and answer business-critical questions on the spot.
- **Be prepared for massive scale**—Infinitely and easily scale your SQL analytics solution by adding an unlimited number of industry-standard servers or choose the pre-configured HPE AppSystem for Vertica, preparing your organization for success in the face of an onslaught of Big Data.
- **Minimize implementation and administration costs**—With built-in simplicity and support for industry standards, the HPE Vertica Analytics Platform can be installed and set up in days (not weeks or months) with fewer costly administration resources, less retraining, and budget-friendly implementation engagements.

Learn how you can store massive volumes of CDR data for a fraction of the cost of traditional databases and data warehouses

Big Data Analytics is an enormous opportunity, particularly for Communication Services Providers (CSPs) to create the intelligence for operating networks more efficiently, analyze the success of services, and create a better personal experience for customers in minimizing churn and increasing market share. However, enterprise data warehouses are exorbitantly cost prohibitive, preventing the storing and analyzing of Big Data in meeting these critical business objectives.

Contact your HPE account manager and learn how you can store massive volumes of CDR data for a fraction of the cost of traditional databases and data warehouses.

Learn more at hpe.com/vertica



Sign up for updates

★ Rate this document



© Copyright 2013, 2016 Hewlett Packard Enterprise Development LP. The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

Oracle is a registered trademark of Oracle and/or its affiliates.

4AA4-7551ENW, March 2016, Rev. 1



Hewlett Packard Enterprise

hpe.com/vertica

HPE Vertica provides solutions to big data challenges. The HPE Vertica Analytics Platform was purpose-built for advanced analytics against big data. It consists of a massively parallel, columnar database as part of the HPE Haven analytics platform. HPE Vertica is known for high performance with very complex analytic queries against multi-terabyte datasets.

HPE Vertica for SQL on Hadoop is a new offering that provides a high-performance, enterprise-ready way to perform ANSI-standard SQL queries on Hadoop data. It integrates with any Hadoop distribution, but HPE has partnered with the major distribution providers—Hortonworks, Cloudera, and MapR—to assure consistent performance and standards across all distributions.

While SQL is the primary query and analysis language, HPE Vertica also supports Java, Python, R, and C. Furthermore, the HPE Vertica Flex Zone feature enables users to define and apply schema during query and analysis, thereby handling exotic data types that are unstructured or schema-free. To simplify and accelerate the deployment of an analytic solution, HPE offers reference architectures and hardware that is optimized for HPE Vertica, although special hardware is not required. Cloud-based versions of the Haven platform are also offered.

[Get a free trial of HPE Vertica today](#)

[For more information, please visit
www.hpe.com/vertica](#)



Advancing all things data.

tdwi.org

TDWI is your source for in-depth education and research on all things data. For 20 years, TDWI has been helping data professionals get smarter so the companies they work for can innovate and grow faster.

TDWI provides individuals and teams with a comprehensive portfolio of business and technical education and research to acquire the knowledge and skills they need, when and where they need them. The in-depth, best-practices-based information TDWI offers can be quickly applied to develop world-class talent across your organization's business and IT functions to enhance analytical, data-driven decision making and performance.

TDWI advances the art and science of realizing business value from data by providing an objective forum where industry experts, solution providers, and practitioners can explore and enhance data competencies, practices, and technologies.

TDWI offers major conferences, topical seminars, onsite education, a worldwide membership program, business intelligence certification, live webinars, resourceful publications, industry news, an in-depth research program, and a comprehensive website: tdwi.org.

© 2015 by TDWI, a division of 1105 Media, Inc. All rights reserved.
Reproductions in whole or in part are prohibited except by written permission.
E-mail requests or feedback to info@tdwi.org.

Product and company names mentioned herein may be trademarks and/or registered trademarks of their respective companies.